

IMI Seminar Series

Spring 2018



Clustering high dimensional data with sparsity

Dominique Picard

Laboratoire de Probabilités et Modèles Aléatoires,
Université Paris VII

MONDAY

April

23

4:00 - 5:00 PM

LeConte College

Room 312

We begin by showing the practical example of finding homogeneous climatological regions in a particular geographical domain (France metropolitan territory). This example, is interesting by itself and shows at the same time the forces and weaknesses of the classification methods. It allows for instance to pose several important theoretical questions.

Among these, we propose to investigate the smoothing problem in the case of high dimensional data. In particular, we try to give answers to the following questions.

- (1) For clustering high dimensional data, what is better: keeping the raw data or smoothing? What does smoothing means?
- (2) What conditions are relevant? in terms of sparsity of the data, in terms of separation of the clusters...
- (3) How to smooth? Does usual adaptation methods work as well to detect clusters?
- (4) Does on-line (signal by signal smoothing) perform as well as off-line smoothing (using a pre-process involving all the signals)?
- (5) What are the rates of convergence?

For more information about this and other IMI seminars, please visit:
<http://imi.cas.sc.edu/events/802/>

